

# High resolution SAR-image classification by Markov random fields and finite mixtures

Gabriele Moser<sup>1a</sup>, Vladimir Krylov<sup>b,c</sup>, Sebastiano B. Serpico<sup>a</sup>, Josiane Zerubia<sup>c</sup>

<sup>a</sup> Dept. of Biophysical and Electronic Engineering (DIBE), University of Genoa,  
Via Opera Pia 11a, I-16145, Genoa (Italy);

<sup>b</sup> Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University,  
119991 Leninskie Gory, Moscow (Russia);

<sup>c</sup> EPI Ariana, CR INRIA Sophia Antipolis Méditerranée, 2004, Route des Lucioles,  
B.P.93, FR-06902, Sophia Antipolis (France).

*Copyright 2010 by SPIE and IS&T. This paper was published in the proceedings of IS&T/SPIE Electronic Imaging 2010 Conference in San Jose, USA, and is made available as an electronic reprint with permission of SPIE and IS&T. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.*

## ABSTRACT

In this paper we develop a novel classification approach for high and very high resolution polarimetric synthetic aperture radar (SAR) amplitude images. This approach combines the Markov random field model to Bayesian image classification and a finite mixture technique for probability density function estimation. The finite mixture modeling is done via a recently proposed dictionary-based stochastic expectation maximization approach for SAR amplitude probability density function estimation. For modeling the joint distribution from marginals corresponding to single polarimetric channels we employ copulas. The accuracy of the developed semiautomatic supervised algorithm is validated in the application of wet soil classification on several high resolution SAR images acquired by TerraSAR-X and COSMO-SkyMed.

**Keywords:** SAR image classification, dictionary, amplitude probability density, stochastic expectation maximization, Markov random field, copula

## 1. INTRODUCTION

In modern remote sensing the use of Synthetic Aperture Radar (SAR) represents an important source of information for Earth observation. SAR images are becoming widely used in various applications, e.g., in flood/fire monitoring, epidemiological surveillance. One of the basic problems arising in most applications is image classification. The classification maps can either be directly used in applications or serve as input to further SAR processing problems like segmentation or change detection.

Contemporary SAR systems are capable of providing fully polarized images (HH/HV/VH/VV). Polarimetric SAR imagery, in comparison to single channel SAR data, have the advantage of a more complete description of landcover's scattering behavior. The possible gain in classification accuracy with data in several polarizations

---

<sup>1</sup>e-mail: G. Moser - gemini@unige.it

compared to one channel data explains the special interest to multi-polarized image classification. In this work we investigate the Dual-polarization SAR case (which can be easily extended to multi-polarization case), as well as single-polarization SAR as a special case.

During the past years, various methods were employed for classification of polarimetric SAR data, based on the maximum likelihood [1], artificial neural networks [2][3], support vector machines [4], fuzzy methods [5] or other approaches [6][7]. In this paper we develop a semiautomatic algorithm for SAR classification using solely the amplitude data and not the complex data as the approaches above. This is considered as an important classification framework both because several image products provided by novel high resolution satellite SAR systems are geocoded ellipsoid-corrected amplitude (intensity) images and because this modality was the main one for several earlier coarser resolution sensors (e.g., ERS). The proposed algorithm combines the Markov random field (MRF) approach to Bayesian image classification with the dictionary-based stochastic expectation maximization (DSEM) amplitude histogram estimator. MRFs represent a general family of probabilistic image models that provide a convenient and consistent way to characterize context dependent data [8]. DSEM was developed for the purpose of modeling the distribution of amplitudes of SAR image. Contrary to the single probability density function (pdf) modeling methods proposed earlier [9][10][11][12][13], it uses a mixture of several distinct pdfs to accurately model the amplitude statistics, thus extending the modeling capacity to heterogeneous images. This makes the developed classification algorithm robust with respect to possibly complicated shapes of class histograms. The DSEM-MRF technique is extended to polarimetric SAR images by modeling the joint distributions of single channels via copulas, resulting in Copula-DSEM-MRF approach (CoDSEM). Copulas [14] is a rapidly developing instrument in statistics designed for constructing joint distributions from marginals with a wide variety of allowed dependence structures. Thus CoDSEM algorithm has good modeling properties based on two flexible statistical modeling concepts.

The paper is organized as follows. In Section 2.1 we present the developed DSEM and CoDSEM approaches for high resolution SAR classification, Sections 2.2 and 2.3 provide overviews of the DSEM approach for amplitude distribution modeling and bivariate copula theory. Section 3 reports experiments on high resolution Dual-Pol TerraSAR-X and Single-Pol COSMO-SkyMed images. Conclusions are drawn in Section 4.

## 2. COPULA-DSEM-MRF APPROACH FOR SAR CLASSIFICATION

### 2.1 Method overview

For the sake of simplicity we present the classification approach in case of 2 polarizations, however the proposed method can easily be extended to multi-polarized case. Consider the Dual-Pol amplitude SAR image and the classification into  $M$  classes.

**DSEM-step.** We begin by estimating class distributions from the observed data from a learning image with ground truth. For every class the mixture pdfs  $p_{dm}(y|\omega_m)$  and cumulative distribution functions (CDFs)  $F_{dm}(y_d|\omega_m)$ , where  $\omega_m$  is the  $m$ -th class, are estimated by DSEM

$$p_{dm}(y_d|\omega_m) = \sum_{i=1}^K P_{dmi} p_{dmi}(y_d|\theta_{dmi}), \quad F_{dm}(y_d|\omega_m) = \sum_{i=1}^K P_{dmi} F_{dmi}(y_d|\theta_{dmi}), \quad (1)$$

with  $m = 1, \dots, M$ ,  $d = 1, 2$ . The use of DSEM instead of single parametric pdf models proves to be advantageous when modeling inhomogeneous classes (i.e. the classes which contain several different landcover

subclasses). However, even in case of homogenous classes, DSEM can be viewed as an efficient tool for choosing the best single pdf model from the panel of pdfs in the DSEM dictionary.

**Copula-step.** The joint distributions  $p_m(\mathbf{y}|\omega_m)$  for classes  $m = 1, \dots, M$  are modeled then via copulas from marginal distributions (1). Such a choice for joint distribution estimation is attractive because of a wide variety of dependence structures provided by copulas. Thus joint pdfs are constructed as

$$p_m(\mathbf{y}|\omega_m) = p_{1m}(y_1|\omega_m)p_{2m}(y_2|\omega_m)\frac{\partial^2 C_m^*}{\partial y_1 \partial y_2}(F_{1m}(y_1|\omega_m), F_{2m}(y_2|\omega_m)), \quad (2)$$

where copula  $C_m^*$  has been chosen for the  $m$ -th class (see Sec. 2.3).

In this work we employed the DSEM dictionary with 4 pdfs instead of the 8 ones used in [15]. This was done in order to facilitate the calculation of (2): for all pdfs in the considered dictionary there are analytical expressions for CDFs, whereas the rest of pdfs in [15] have no analytical expression for  $F_{dm}(y_d)$  and the numerical computations are costly. In addition, tests with a full dictionary of 8 pdfs reported a minor improvement in modeling accuracy (within 1% of correlation  $\rho$  increase for the same number of components) compared to the considered dictionary of 4.

**MRF-step.** In order to gain robustness against speckle, we adopt a contextual approach based on a Markov Random Field (MRF) model. We skip here the classical definition of MRFs and hidden MRFs (see e.g., [16][17]). Introducing an anisotropic second-order neighborhood system with only cliques  $C$  of size 2 considered, we get the following MRF energy function with parameter  $\beta > 0$  for the class labels:

$$H(x|\beta) = \sum_c V(x_c|\beta) = \sum_{c=\{s,s'\} \in C} [-\beta \delta_{x_s=x_{s'}}], \quad \text{with } \delta_{x_s=x_{s'}} = \begin{cases} 1, & \text{if } x_s = x_{s'} \\ 0, & \text{otherwise} \end{cases}.$$

As in a usual hidden MRF model, the unobserved class labels are modeled as an MRF and observations are conditionally independent given labels [16]. Given the conditional pdfs (2) and no prior information about the proportions of classes on the testing image (which results in giving the same prior weights to all the classes) the resulting energy function is:

$$U(\omega_m|\mathbf{y}, \beta) = \sum_{i \in S} \left[ -\log p_m(\mathbf{y}|\omega_m) - \beta \sum_{s:\{i,s\} \in C} \delta_{x_i=x_s} \right]. \quad (3)$$

**Estimation-step.** As can be seen from the energy function specified in (1)-(3) there is only 1 parameter  $\beta$  to estimate. This is thanks to the fact that all the parameter estimation for (1) is incorporated into DSEM.

For  $\beta$  estimation we suggest to use simulated annealing [18] with pseudo-likelihood function  $PL$  [17]:

$$\log PL(x|\beta) = \log \left[ \prod_{s \in S} p(x_s|x_{S \setminus \{s\}}, \beta) \right], \quad \text{where } p(x_s|x_{S \setminus \{s\}}, \beta) = \frac{\exp(-U(x_s|x_{S \setminus \{s\}}, \beta))}{\sum_{z_s \in X_S} \exp(-U(z_s|x_{S \setminus \{s\}}, \beta))},$$

with normal proposal distribution  $N(\beta_t, 1)$  and cooling schedule  $T_t = 0.95 T_{t-1}$ . If necessary, this procedure can be followed by the method suggested in [19], combining Metropolis-Hastings algorithm with gradient method. The final estimate  $\beta^*$  is chosen by averaging the estimates on the last  $t$  iterations  $\beta^* = \left( \sum_{i=N-t}^N \beta_i \right) / t$ .

**Optimization-step.** By this time the CoDSEM-MRF model is completely specified. This step involves the minimization of the energy (3). For this optimization the deterministic Modified Metropolis Dynamics (MMD) [20] algorithm is suggested. It is a compromise solution between the deterministic Iterated Conditional Modes algorithm [17], which is fast and finds local a minimum, and Simulated Annealing [18], that is slow, but finds the global minimum. MMD algorithm proceeds as follows:

1. sample a random initial configuration  $\omega^0$ , with  $k = 0$  and  $T = T_0$ ;
2. using uniform distribution pick up a global state  $\eta$  which differs exactly in one element from  $\omega^k$ ;
3. compute  $\Delta U = U(\eta) - U(\omega)$  and accept  $\eta$  according to the rule:

$$\omega^{k+1} = \begin{cases} \eta, & \text{if } \Delta U \leq 0, \\ \eta, & \text{if } \Delta U \geq 0 \text{ and } \ln(\alpha) \leq -\frac{\Delta U}{T}, \\ \omega^k, & \text{otherwise.} \end{cases}$$

where  $\alpha$  is a **constant** threshold ( $\alpha \in (0, 1)$ ), chosen at the start of the algorithm;

4. decrease the temperature  $T = T_{k+1}$  and goto Step 2 until convergence ( $\Delta U/U$  below a chosen threshold).

MMD demonstrated good results for the minimization problem (1)-(3).

In case of a single channel polarization there is no need for the **Copula-step**, because the **DSEM-step** estimates can be directly plugged into (3), thus in single-pol case the DSEM-classification algorithm goes:

**DSEM-step  $\rightarrow$  MRF-step  $\rightarrow$  Estimation-step  $\rightarrow$  Optimization-step.**

## 2.2 Dictionary-based Stochastic Expectation Maximization

The Dictionary-based Stochastic Expectation Maximization (DSEM) approach was proposed in [21], then developed and validated on high-resolution SAR in [15]. It models a SAR-image amplitude distribution as a finite mixture model (FMM) with components taken from a dictionary of SAR-specific pdfs. DSEM integrates the number of mixture component estimation, selection of the optimal model for each component from a dictionary  $\mathcal{D}$ , and parameter estimation for mixture components.

To formalize the scenario when several distinct land-cover typologies are present in the same SAR image, a FMM [22] for the distribution of grey levels is assumed. The amplitude SAR image is modeled as a set  $\mathcal{I} = \{r_1, \dots, r_N\}$  of independent and identically distributed samples drawn from the mixture pdf:

$$p_r(r|\theta) = \sum_{i=1}^K P_i f_i(r|\theta_i), \quad r \geq 0, \quad (4)$$

where  $f_i(r|\theta_i)$  are pdfs with parameters  $\theta_i$ , taking values in a set  $\Theta_i \subset \mathbb{R}^{\ell_i}$  and  $\{P_i\}$  are mixing proportions:  $\sum_{i=1}^K P_i = 1$  with  $0 \leq P_i \leq 1$ ,  $i = 1, \dots, K$ . Following [21][15][16], each component  $f_i(r|\theta_i)$  in (4) is modeled by resorting to a finite dictionary  $\mathcal{D} = \{f_1, f_2, f_3, f_4\}$  (see Table 1) of 4 SAR-specific distinct parametric pdfs  $f_i(r|\theta_i)$ , parameterized by  $\theta_i \in A_i$ ,  $i = 1, \dots, 4$ . For descriptions of distributions in  $\mathcal{D}$  and their physical properties see [23].

As discussed in [21], considering the variety of estimation approaches for FMMs the choice for this particular estimation problem is the stochastic expectation maximization (SEM) scheme [25]. SEM is an iterative estimation procedure dealing with the problem of *data incompleteness*. In case of FMMs the complete data is

Table 1: Pdfs and MoLC equations for the parametric families included in the considered dictionary  $\mathcal{D}$ . Here  $\Gamma(\cdot)$  is the Gamma function [24],  $\Psi(\cdot)$  the Digamma function [24] and  $\Psi(\nu, \cdot)$  the  $\nu$ th order polygamma function [24].

Family	Probability density function	MoLC equations
Log-normal	$f_1(r) = \frac{1}{\sigma r \sqrt{2\pi}} \exp \left[ -\frac{(\ln r - m)^2}{2\sigma^2} \right]$	$\kappa_1 = m$ $\kappa_2 = \sigma^2$
Weibull	$f_2(r) = \frac{\eta}{\mu^\eta} r^{\eta-1} \exp \left[ -\left(\frac{r}{\mu}\right)^\eta \right]$	$\kappa_1 = \ln \mu + \Psi(1)\eta^{-1}$ $\kappa_2 = \Psi(1, 1)\eta^{-2}$
Nakagami	$f_3(r) = \frac{2}{\Gamma(L)} (\lambda L)^L r^{2L-1} \exp(-\lambda L r^2)$	$2\kappa_1 = \Psi(L) - \ln \lambda L$ $4\kappa_2 = \Psi(1, L)$
Generalized Gamma	$f_4(r) = \frac{\nu}{\sigma \Gamma(\kappa)} \left(\frac{r}{\sigma}\right)^{\kappa\nu-1} \exp \left\{ -\left(\frac{r}{\sigma}\right)^\nu \right\}$	$\kappa_1 = \Psi(\kappa)/\nu + \ln \sigma$ $\kappa_2 = \Psi(1, \kappa)/\nu^2$ $\kappa_3 = \Psi(2, \kappa)/\nu^3$

represented by the set  $\{(r_i, s_i), i = 1, \dots, N\}$ , where  $r_i$  are the observations (SAR amplitudes) and  $s_i$  are the missing labels: given an FMM with  $K$  components,  $s_i$  takes value in  $\{1, \dots, K\}$  and denotes to which of the  $K$  components the  $i$ -th pixel belongs.

Instead of adopting ML estimates as in the classical SEM scheme [25], in DSEM the Method of Log-Cumulants (MoLC) is adopted [10][21], which has been demonstrated to be a feasible and effective estimation tool for common SAR parametric models [10] and also for all pdfs in  $\mathcal{D}$  [15]. MoLC has recently been proposed as a parametric pdf estimation technique suitable for distributions defined on  $[0, +\infty)$ , it has been widely applied in the context of SAR-specific parametric families for amplitude and intensity data modeling, e.g., for Nakagami distribution [10]. MoLC adopts the Mellin transform [24] by analogy to the Laplace transform in moment generating function [26], and allows stating a set of equations relating the unknown parameters of a given parametric model with one or several log-cumulants:

$$\kappa_1 = E\{\ln u\}, \quad \kappa_2 = \text{Var}\{\ln u\}, \quad \kappa_3 = E\{(\ln u - \kappa_1)^3\}.$$

These equations have one solution for any observed log-cumulants for all pdfs in  $\mathcal{D}$  (see Table 1). A refined procedure to estimate the number of mixture components was integrated in DSEM in [15].

Each iteration of DSEM goes as follows:

- **E-step:** compute, for each greylevel  $z$  and  $i$ -th component, the posterior probability estimates corresponding to the current pdf estimates, i.e.  $z = 0, \dots, Z-1$ ,  $i = 1, \dots, K_t$ :

$$\tau_i^t(z) = \frac{P_i^t p_i^t(z)}{\sum_{j=1}^{K_t} P_j^t p_j^t(z)};$$

- **S-step:** sample the label  $s^t(z)$  of each greylevel  $z$  according to the current estimated posterior probability distribution  $\{\tau_i^t(z) : i = 1, \dots, K_t\}$ ,  $z = 0, \dots, Z-1$ ;
- **MoLC-step:** for the  $i$ -th mixture component, compute the following histogram-based estimates of the mixture proportion and the first three log-cumulants:

$$P_i^{t+1} = \frac{\sum_{z \in Q_{it}} h(z)}{\sum_{z=0}^{Z-1} h(z)}, \quad \kappa_{1i}^t = \frac{\sum_{z \in Q_{it}} h(z) \ln z}{\sum_{z \in Q_{it}} h(z)}, \quad \kappa_{bi}^t = \frac{\sum_{z \in Q_{it}} h(z) (\ln z - \kappa_{1i}^t)^b}{\sum_{z \in Q_{it}} h(z)},$$

where  $b = 2, 3$ ,  $h(z)$  is the image histogram,  $Q_{it} = \{z : s^t(z) = \sigma_i\}$  is the set of grey levels assigned to the  $i$ -th component; then, solve the corresponding MoLC equations (see Table 1) for each parametric family  $f_j(\cdot|\theta_j)$  ( $\theta_j \in A_j$ ) in the dictionary, thus computing the resulting MoLC estimate  $\theta_{ij}^t$ ,  $i = 1, \dots, K_t$ ,  $j = 1, \dots, M$ ;

- **K-step:** for all  $i$ ,  $i = 1, \dots, K_t$ : if  $P_i^{t+1}$  is below a given threshold, eliminate the  $i$ -th component, update  $K_{t+1}$ ;
- **Model Selection-step:** for the  $i$ -th mixture component, compute the log-likelihood of each estimated pdf  $f_j(\cdot|\theta_{ij}^t)$  according to the data assigned to the  $i$ -th component:

$$L_{ij}^t = \sum_{z \in Q_{it}} h(z) \ln f_j(z|\theta_{ij}^t)$$

and define  $p_i^{t+1}(\cdot)$  as the estimated pdf  $f_j(\cdot|\theta_{ij}^t)$  yielding the highest value of  $L_{ij}^t$ ,  $i = 1, \dots, K_{t+1}$ ,  $j = 1, \dots, 4$ .

### 2.3 Copulas

For the purpose of modeling the joint distribution of a multivariate random vectors given marginal distributions we employ copulas [14]. It is worth nothing that many current SAR sensors include a Dual-Pol acquisition modality; this further enforces the interest in the classification in two-dimensional SAR amplitude/intensity spaces. Indeed, as mentioned earlier, many parametric models have been proposed for the marginal statistics of SAR amplitudes or intensities, but only very few models are available for the joint distribution of several SAR amplitudes (e.g., Nakagami-Gamma model developed in [27]). In order to overcome this limitation, we combine the marginal pdfs provided by DSEM with copulas. Here we present a brief overview of copula concepts. For the sake of simplicity we present only the 2D case of copulas, though all the definitions and properties presented can be extended to multivariate cases.

2D copula is a bivariate joint distribution defined on  $[0, 1]^2$  such that marginal distributions are uniform on  $[0, 1]$ . Specifically, A *bivariate copula* is a function  $C : [0, 1]^2 \rightarrow [0, 1]$ , which satisfies the following properties:

1. both marginals are uniformly distributed on  $[0, 1]$ ;
2. for every  $u, v$  in  $[0, 1]$ :  $C(u, 0) = C(0, v) = 0$ , and  $C(u, 1) = u$ ,  $C(1, v) = v$ ;
3. for every  $u_1 \leq u_2$ ,  $v_1 \leq v_2$  in  $[0, 1]$ :  $C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$ .

The importance of copulas in statistics is explained by Sklar's Theorem [14], which states the existence of a copula  $C$ , that models the joint distribution function  $H$  of arbitrary r.v.s  $X$  and  $Y$  with CDFs  $F$  and  $G$ :

$$H(x, y) = C(F(x), G(y)), \quad (5)$$

for all  $x, y$  in  $\mathbb{R}$ , and if  $F$  and  $G$  are continuous, then  $C$  is unique. Thus copulas link joint distribution functions to their one-dimensional marginals.

Given absolutely continuous random variables with pdfs  $f(x)$  and  $g(y)$  and corresponding CDFs  $F(x)$  and  $G(y)$ , the pdf of the joint distribution  $h(x, y)$  corresponding to (5) is given by:

$$h(x, y) = f(x)g(y) \frac{\partial^2 C}{\partial x \partial y}(F(x), G(y)), \quad (6)$$

where  $\frac{\partial^2 C}{\partial x \partial y}(x, y)$  is pdf corresponding to the copula  $C(x, y)$ .

An important family of copulas are archimedean copulas, which have a simple analytical form and yet exhibiting a wide variety of modeled dependence structures.

An *archimedean* copula is a bivariate copula  $C$ , defined as  $C(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2))$ , where the *generator function*  $\phi(u)$  is any function satisfying the following properties: 1.  $\phi(u)$  is continuous on  $[0, 1]$ ; 2.  $\phi(u)$  is decreasing,  $\phi(1) = 0$ ; 3.  $\phi(u)$  is convex.

Here we consider 5 explicitly defined copulas: 4 archimedean (Clayton, Ali-Mikhail-Haq, Gumbel, Frank) and 1 non-archimedean copula with simultaneous presence of an absolutely continuous and a singular component (Marchal-Olkin) [14]. This choice of copulas is capable of modeling a considerable variety of dependence structure [28]. We summarize the information about the copulas used in this work in Table 2.

Table 2: Considered copulas: Clayton, Ali-Mikhail-Haq (AMH), Gumbel, Frank and Marchal-Olkin (Marchal), defined by their  $C(u, v)$ , with  $\theta(\tau)$  and  $\tau$ -intervals.

Copula	$C(u, v)$	$\theta(\tau)$	$\tau$ interval
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$\theta = \frac{2\tau}{1-\tau}$	$\tau \in (0, 1]$
AMH	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\tau = \frac{3\theta-2}{3\theta} - \frac{2}{3} \left(1 - \frac{1}{\theta}\right)^2 \log(1-\theta)$	$\tau \in [-0.181726, \frac{1}{3}]$
Gumbel	$\exp\left(-\left[(-\log(u))^\theta + (-\log(v))^\theta\right]^{1/\theta}\right)$	$\theta = \frac{1}{1-\tau}$	$\tau \in [0, 1]$
Frank	$-\frac{1}{\theta} \log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$\tau = 1 - \frac{4}{\theta^2} \int_0^\theta \frac{t}{e^{-t}-1} dt$	$\tau \in [-1, 0) \cup (0, 1]$
Marchal	$\min(u^{1-\theta}v, uv^{1-\theta})$	$\theta = \frac{2\tau}{\tau+1}$	$\tau \in [0, 1]$

A common way of copula estimation is by using its connection with Kendall's tau, which is a ranking correlation coefficient [14]. *Kendall's tau* is a concordance-discordance measure between two independent realizations  $(Z_1, Z_2)$  and  $(\hat{Z}_1, \hat{Z}_2)$  from the same law  $H(x, y)$ :  $\tau = P\{(Z_1 - \hat{Z}_1)(Z_2 - \hat{Z}_2) > 0\} - P\{(Z_1 - \hat{Z}_1)(Z_2 - \hat{Z}_2) < 0\}$ . Given two realizations  $z_{1,l}$  and  $z_{2,l}$  ( $l = 1, \dots, N$ ), the empirical estimator of Kendall's tau is

$$\hat{\tau} = \frac{\sum_{l=1}^{N-1} \sum_{k=l+1}^N z_{1,lk} z_{2,lk}}{\binom{N}{2}}, \quad \text{where } z_{n,lk} = \begin{cases} 1, & \text{if } z_{n,l} \leq z_{n,k} \\ -1, & \text{otherwise} \end{cases}, \quad \text{for } n = 1, 2. \quad (7)$$

By integrating in the definition of  $\tau$  over the distribution of  $(\hat{Z}_1, \hat{Z}_2)$ , we get

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1, \quad \text{and} \quad \tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt, \quad (8)$$

where the first equation gives the general connection between Kendall's  $\tau$  and the copula  $C$  associated with  $H(x, y)$ , and the second gives the connection with archimedean copulas in terms of generator function  $\phi(t)$ . All copulas considered in this work are defined by a single parameter  $\theta$ . By plugging the empirical estimate  $\hat{\tau}$  (7) in place of  $\tau$  in (8), we get parameter estimates  $\hat{\theta}$  (see Table 2).

Moreover, given an estimator of tau  $\hat{\tau}$  one should decide whether the specific copula under consideration is appropriate for modeling the dependence of this level  $\hat{\tau}$ . Some copulas are specific to marginals with low level of

correlation, others deal with strongly correlated marginals and other ones are capable of modeling all levels of rank correlation. So in every case the list of copulas is limited to those which are capable of accurately modeling the specific value  $\hat{\tau}$ .

Now we address the problem of copula choice (from listed in Table 2). First we discard those copulas, for which the current estimate  $\hat{\tau}_m$  is outside the  $\tau$ -relevance interval (see Table 2). Then, for each  $m$ , we choose the best fitting copula, i.e. the copula with the highest  $p$ -value in Pearson Chi-square test-of-fitness (PCS) [29]. The null hypothesis in PCS is that the sample frequencies  $C_c(F_{1m}(u_1), F_{2m}(u_2))$  ( $m = 1, \dots, M$ ,  $c = 1, \dots, 5$ ), where  $(u_1, u_2)$  are the observed data, are consistent with the theoretical frequencies for the copula  $c$ :  $C_c(v_1, v_2)$ .

### 3. EXPERIMENTS

In this section we present experiments with DSEM and CoDSEM classification algorithms in application to flooded/wet/dry soil mapping for epidemiology monitoring. Thus  $M = 3$  classes are: water, wet, and dry soil.

We present here tests with 3 SAR images:

- **TSX1, TSX2:** Dual HH/VV Polarized **TerraSAR-X**, Stripmap (6 m ground resolution), geocoded, 2-look image acquired over China, Sanchagang; (©Infoterra); see Fig. 1, Fig. 2.
- **CSK1:** Single HH Polarized **COSMO-SkyMed**, Stripmap (2.5 m ground resolution), geocoded, single-look image acquired over Italy, Piemonte; (©ASI); see Fig. 3.

For all images, a small portion (around 500x500 pixels) was used for learning on manually created ground truth (GT) maps. The rest of each image was used for testing (**Optimization-step**) and the resulting classification maps and confusion matrices are reported.

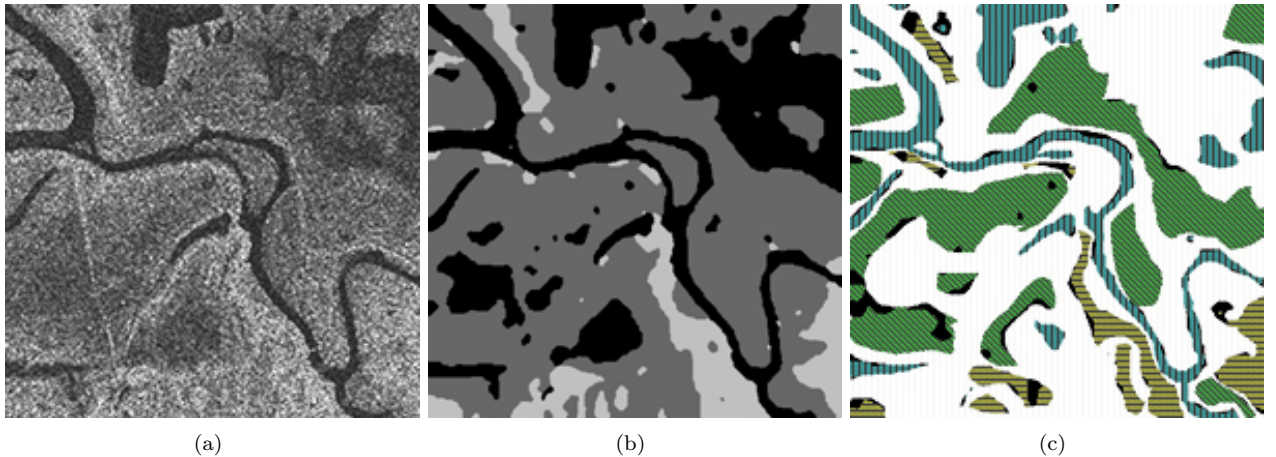


Figure 1: **TSX2**. HH polarization (a), full CoDSEM-classification map (b) (black - water, grey - wet soil, white - dry soil), and map of correct classification (c) referenced to nonexhaustive GT (white - outside GT, correctly classified water "✓", wet soil "✓", dry soil "✓", black - misclassification of all types) (TerraSAR-X, ©Infoterra).



Table 3: Accuracy and confusion matrices for CoDSEM, 2D Nakagami-Gamma and  $K$ -NN classification on **TSX1**.

Image	TSX1 (CoDSEM)			TSX1 (2D Nakagami-Gamma)			TSX1 ( $K$ -NN)		
Overall	97.07%			90.22%			95.92%		
Class	water	wet	dry	water	wet	dry	water	wet	dry
water	<b>99.37%</b>	0.21%	0.40%	<b>98.75%</b>	0.92%	0.28%	<b>98.79%</b>	0.83%	0.36%
wet	1.00%	<b>96.83%</b>	2.15%	1.82%	<b>94.63%</b>	3.51%	0.25%	<b>98.01%</b>	1.74%
dry	0.09%	4.17%	<b>95.73%</b>	0.09%	24.33%	<b>75.50%</b>	0.12%	9.07%	<b>90.80%</b>

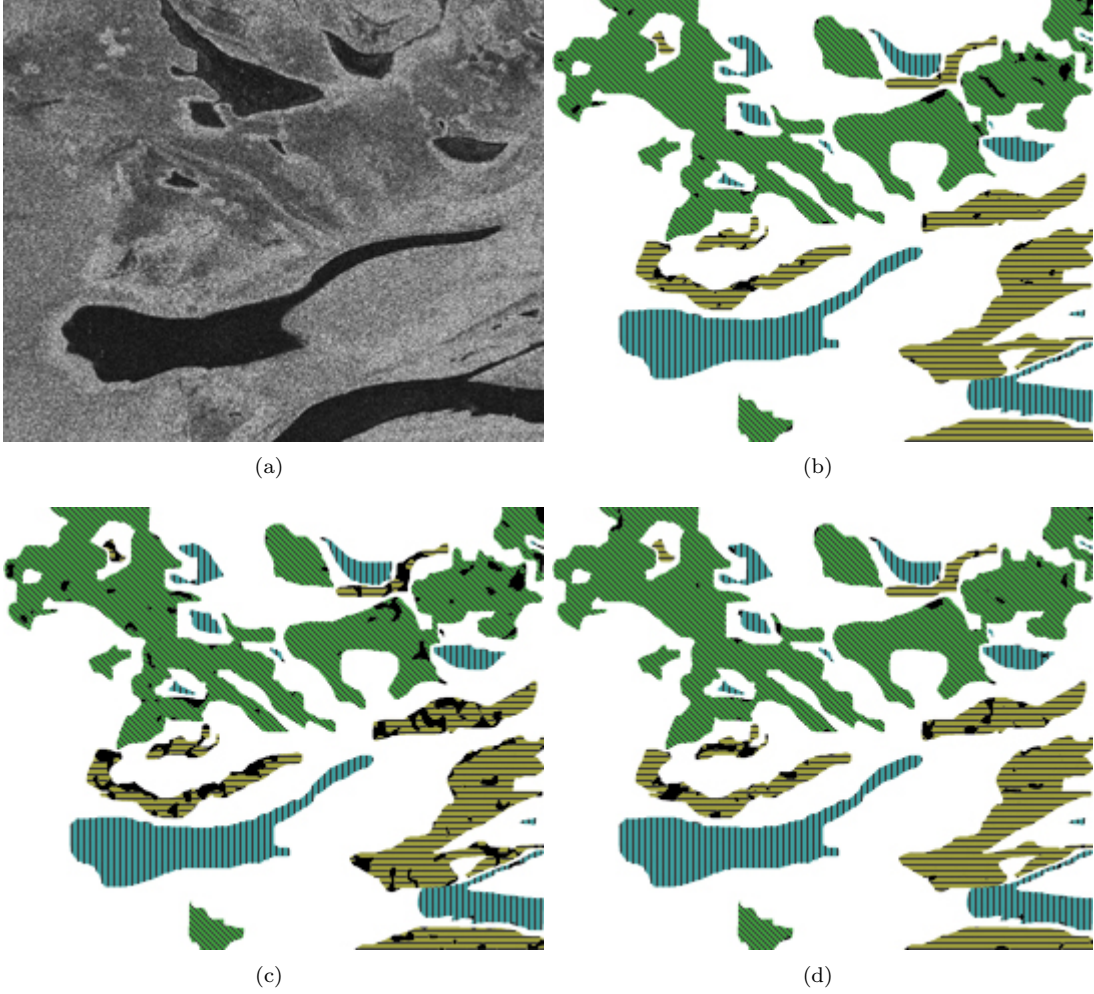


Figure 2: **TSX1**. HH polarization (a), CoDSEM-classification map (b) referenced to nonexhaustive GT (white - outside GT, correctly classified water "|", wet soil "\", dry soil "-"), 2D Nakagami-Gamma classification (c), and  $K$ -NN classification (d) (TerraSAR-X, ©Infoterra).

Here are some comments to the reported experiments:

- **TSX1.** CoDSEM reported an overall accuracy of 97% (see Table 3). For the sake of comparison, we provide classification maps (Fig. 2) and accuracy (see Table 3) acquired by 2D Nakagami-Gamma model [27] (Fig. 2(c)) and  $K$ -nearest neighbors [30] (Fig. 2(d)), with  $K^* = 60$  estimated by cross-validation. The respective models were combined with MRF approach, same estimation and optimization procedures were used as described in Sec. 2.1. We stress here the higher classification accuracy achieved by CoDSEM as compared to both parametric density estimation 2D Nakagami-Gamma approach and nonparametric general classification approach  $K$ -NN.
- **TSX2.** CoDSEM reported an overall accuracy of 92% (see Table 4). For this image we present the full classification map (Fig. 1(b)) and the map of correct classification referenced to nonexhaustive GT (Fig. 1(c)).
- **CSK1.** This image has only HH polarization, thus DSEM classification was used, reporting an overall accuracy of 91% (see Table 4). This accuracy is somewhat inferior to those of CoDSEM, as it operates on a single polarization. More exhaustive testing of DSEM and CoDSEM [16] revealed an average gain of 5 – 10% in total classification accuracy from adding the second polarization channel.

Table 4: Accuracy and confusion matrices for the proposed algorithm on **TSX2** and **CSK1**.

Image	TSX2 (CoDSEM)			CSK1 (DSEM)		
Overall	92.00%			91.61%		
Class	water	wet	dry	water	wet	dry
water	93.12%	6.32%	0.54%	86.20%	13.55%	0.23%
wet	5.97%	93.76%	0.26%	0.76%	92.39%	6.84%
dry	0.03%	15.54%	84.42%	0.03%	4.66%	95.29%

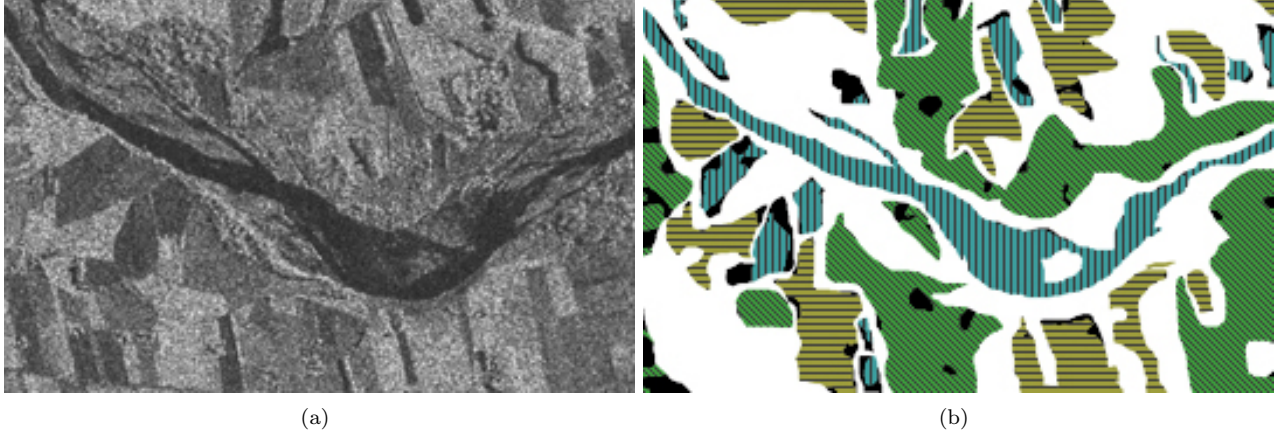


Figure 3: **CSK1**. HH polarization (a), DSEM-classification map (b) referenced to nonexhaustive GT (white - outside GT, correctly classified water "G", wet soil "B", dry soil "Y", black - misclassification of all types) (COSMO-SkyMed, ©ASI).

## 4. CONCLUSIONS

A novel classification algorithm for high and very high resolution SAR is proposed. It combines the Markov random field approach to Bayesian image classification and a finite mixture technique for probability density function estimation. Finite mixture modeling was done via dictionary-based stochastic expectation maximization approach for amplitude histogram estimation, which provides a good level of accuracy for the algorithm, regardless of the possibly complicated histograms for some classes. The developed algorithm is extended to polarimetric SAR by modeling the joint distributions of channels via copulas. The resulting classification algorithm is supervised (model estimation is done on a training image) and semiautomatic (several DSEM and MMD parameters have to be specified).

The accuracy of the proposed algorithm was validated in the application to flooded/wet/dry soil classification on several high resolution SAR images: Dual-Pol TerraSAR-X and Single-Pol COSMO-SkyMed. The experiments demonstrated a high level of accuracy on all test images.

## ACKNOWLEDGMENT

This work was carried out with the partial financial support of Institut National de Recherche en Informatique et Automatique, France (INRIA) and the French Space Agency (CNES). This support is gratefully acknowledged. The authors would like to thank the Italian Space Agency (ASI) for providing the COSMO-SkyMed image of Piemonte (©ASI, 2008). The TerraSAR-X image of Sanchagang was taken from <http://www.infoterra.de/>, where it is available for free testing (©Infoterra, 2008).

## REFERENCES

- [1] Lee, J.-S., Grunes, M. R., and Kwok, R., “Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution,” *Int. J. Remote Sensing* **15**(11), 2299–2311 (1994).
- [2] Ito, Y. and Omatu, S., “Polarimetric SAR data classification using competitive neural networks,” *Int. J. Remote Sensing* **19**(14), 2665–2684 (1998).
- [3] Chen, K. S., Huang, W. P., Tsay, D. H., and Amar, F., “Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network,” *IEEE Trans. Geosci. Remote Sens.* **34**(3), 814–820 (1996).
- [4] Fukuda, S., Kataqiri, R., and Hirose, H., “Unsupervised approach for polarimetric SAR image classification using support vector machines,” in [*Proceedings of IGARSS*], **5**, 2599–2601 (2002).
- [5] Chen, C. T., Chen, K. S., and Lee, J.-S., “The use of fully polarimetric information for the fuzzy neural classification of SAR images,” *IEEE Trans. Geosci. Remote Sens.* **41**(9), 2089–2100 (2003).
- [6] Cloude, S. R. and Pottier, E., “An entropy based classification scheme for land applications of polarimetric SAR,” *IEEE Trans. Geosci. Remote Sens.* **35**(1), 68–78 (1997).

- [7] Lee, J.-S., Grunes, M. R., Pottier, E., and Ferro-Famil, L., "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sens.* **42**(4), 722–731 (2004).
- [8] Kato, Z., Zerubia, J., and Berthod, M., "Unsupervised parallel image classification using Markovian models," *Pattern Recognition* **32**(4), 591–604 (1999).
- [9] Oliver, C. and Quegan, S., [*Understanding Synthetic Aperture Radar Images*], Artech House, Norwood (1998).
- [10] Tison, C., Nicolas, J.-M., Tupin, F., and Maitre, H., "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.* **42**(10), 2046–2057 (2004).
- [11] Kuruoglu, E. E. and Zerubia, J., "Modelling SAR images with a generalization of the Rayleigh distribution," *IEEE Trans. Image Process.* **13**(4), 527–533 (2004).
- [12] Moser, G., Zerubia, J., and Serpico, S. B., "SAR amplitude probability density function estimation based on a generalized Gaussian model," *IEEE Trans. Image Process.* **15**(6), 1429–1442 (2006).
- [13] Li, H.-C., Hong, W., and Wu, Y.-R., "Generalized Gamma distribution with MoLC estimation for statistical modeling of SAR images," in [*Proceedings of APSAR*], 525–528 (2007).
- [14] Nelsen, R. B., [*An Introduction to Copulas*], Springer, New-York, 2nd ed. (2007).
- [15] Krylov, V., Moser, G., Serpico, S., and Zerubia, J., "Dictionary-based probability density function estimation for high-resolution SAR data," in [*Proceedings of SPIE*], **7246**, 72460S (2009).
- [16] Krylov, V. and Zerubia, J., "High resolution SAR image classification," Research Report 7108, INRIA (2009).
- [17] Besag, J., "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society B* **48**, 259–302 (1986).
- [18] Geman, S. and Geman, D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.* **6**, 721–741 (1984).
- [19] Yu, Y. and Cheng, Q., "MRF parameter estimation by an accelerated method," *Pattern Recognit. Lett.* **24**(9-10), 1251–1259 (2003).
- [20] Kato, Z., Zerubia, J., and Berthod, M., "Satellite image classification using a modified Metropolis dynamics," in [*Proceedings of ICASSP*], 573–576 (1992).
- [21] Moser, G., Serpico, S., and Zerubia, J., "Dictionary-based Stochastic Expectation Maximization for SAR amplitude probability density function estimation," *IEEE Trans. Geosci. Remote Sens.* **44**(1), 188–199 (2006).
- [22] Figueiredo, M. A. F. and Jain, A. K., "Unsupervised learning of finite mixture models," *IEEE Trans. Patt. Anal. Mach. Intell.* **24**(3), 381–396 (2002).
- [23] Krylov, V., Moser, G., Serpico, S., and Zerubia, J., "Modeling the statistics of high resolution SAR images," Research Report 6722, INRIA (2008).

- [24] Sneddon, I., [*The use of integral transforms*], McGraw-Hill, New York (1972).
- [25] Celeux, G., Chauveau, D., and Diebolt, J., “On stochastic versions of the EM algorithm,” Research Report 2514, INRIA (1995).
- [26] Nicolas, J.-M. and Maruani, A., “Lower-order statistics: a new approach for probability density functions defined on  $\mathbb{R}^+$ ,” in [*Proceedings of EUSIPCO*], (2000).
- [27] Lee, J.-S., Hoppel, K. W., Mango, S. A., and Miller, A. R., “Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery,” *IEEE Trans. Geosci. Remote Sens.* **32**(5), 1017–1028 (1994).
- [28] Huard, D., Évin, G., and Favre, A.-C., “Bayesian copula selection,” *Computational Statistics & Data Analysis* **51**(2), 809–822 (2006).
- [29] Lehmann, E. and Romano, J., [*Testing statistical hypotheses*], Springer, New York, 3rd ed. (2005).
- [30] Bishop, C. M., [*Pattern Recognition and Machine Learning*], Springer, New-York (2006).